DOCUMENT RESUME

ED 204 411 TM 810 485

AUTHOR Lazar-Morrison, Charlotte: And Others
TITLE A Review of the Literature on Test Use.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
REPORT NO CSE-R-44
PUB DATE Aug 80
NOTE 32p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Finance: *Elementary Secondary Education: *Instructional Improvement: *Literature Reviews: Minimum Competency Testing: Standardized Tests: Teacher Attitudes: Teacher Education: Teacher Made Tests: *Testing: Test Results
IDENTIFIERS Center for the Study of Evaluation CA: *Test Use

ABSTRACT

This report was developed in the Test Use Project at the Center for the Study of Evaluation. The general goal of the project is to examine and describe the features and applications of tests and other assessment methods that contribute to the improvement of instruction. It is, therefore, examining the nature of current assessment practice, the kinds of information such practice yields, the factors influencing the use of the information, the kinds of uses made, and the costs associated with assessment. The first phase of the project (December 1979 to November 1981) consists primarily of a national survey of teachers and principals on the kind of issues suggested above. The design of the phase 1 national survey has been influenced by a variety of project planning activities. This report deals with one of these activities--a review of the literature on test use. This document integrates the literature reviewed and is intended to be selective rather than exhaustive. The focus is on: (1) the nature of current testing practices: (2) factors influencing the use of test results: and (3) costs associated with testing. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204411

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
[ ] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A REVIEW OF THE LITERATURE ON TEST USE

Charlotte Lazar-Morrison
Linda Polin
Raymond Moy
James Burry

CSE Report No. 144
August 1980

TM 810 485

Test Use Project
CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California . Los Angeles

Table of Contents

	Page
Abstract	1
Preface	2
The Nature of Current Testing Practices	3
Factors Related to Use of Tests and Test Results	9
Teacher training and experience	9
Teacher attitudes	15
Organization of instruction and related variables	19
Costs of Testing	20
Summary	24

Abstract

This report was developed in CSE's Test Use Project. The general goal of the project is to examine and describe the features and applications of tests and other assessment methods that contribute to the improvement of instruction. The Test Use Project is therefore examining the nature of current assessment practice, the kinds of information such practice yields, the factors influencing the use of the information, the kinds of uses made, and the costs associated with assessment.

The first phase of the project (December 1979 to November 1981) consists primarily of a national survey of teachers and principals on the kind of issues suggested above. The second, overlapping phase (February 1981 to November 1982) will consist of a small-scale but intensive study of testing costs and the factors identified in phase 1 as influencing the use of test results.

The design of the phase 1 national survey has been influenced by a variety of project planning activities. This report deals with one of these activities--a review of the literature on test use.

Preface

This document integrates the literature reviewed on test use issues and is designed to help refine the conceptual framework guiding project work and to inform development of the national survey questionnaire and subsequent site visits. It is intended to be a selective review of relevant literature rather than an exhaustive discussion. The major goals are to integrate relevant findings, to move toward broader understanding of test use, and to make recommendations regarding teachers and testing.

The review focuses on three pertinent questions which correspond to and include key components of the project's conceptual framework. The questions are:

(1) What is the nature of current testing practices?

(2) What factors influence the use of test results? and

(3) What costs are associated with testing?

The Test Use Project defines testing broadly to allow the investigation to deal with a full range of assessments or other evaluative procedures and instruments, both formal and informal in nature. Further, this investigation is intended to be from a classroom and/or teacher perspective.

This broad scope of assessment corresponds to what Lortie has described as "tests" and "observations:"

> The monitoring techniques available to teachers are limited in number and precision; essentially, they must rely on various tests of students' knowledge and on observations of how students behave in the classroom. Tests include teacher-prepared examinations, verbal quizzes, student workbooks, and standardized tests. Observing student behavior included judging student interest, watching work effort, checking compliance, and noting the degree of responsiveness to the teacher (Lortie, 1975, p. 138)

The appropriateness of including such formal and informal procedures is supported in empirical studies of sources of teacher information for instructional decision-making (Airasian, 1979; Borko, 1978; Yeh, 1978) as well as in Lortie's own investigation of teaching (Lortie, 1975).

The Nature of Current Testing Practices

There is little research-based information about current testing practice. Almost ten years ago, Kirkland (1971) reviewed the literature on test impact on students and schools and found that while much had been written about tests, few empirical studies were evident. In her review, Kirkland excluded the issues of test validity and misuse. She states that "Many of the issues and related questions are unanswerable by empirical means; for other issues, there is only indirect evidence. An attempt was made to present the current state of the field, rather than implications ... " (p. 307). What is significant about her exclusions is the correct observation that these issues are "implications," often not founded on empirical research. Today, there still remains a plethora of publications on these very issues and a dearth of empirical support on actual test use practices.

Kirkland's review of the literature is concentrated mainly upon the social and psychological issues in testing (e.g., self-concept, anxiety, motivation, level of aspiration), more than upon instructional issues (e.g., curriculum, evaluation and pedagogy). Also, then as now, little empirical research had accumulated on the latter.

Only recently has the testing dialogue begun to move away from social and psychological issues, although not completely, and begun to focus upon the instructional issues of testing. This shift is reflected in the recent

review of literature done by the Institute for Research on Teaching (Rudman, Kelly, Wanous, Mehrens, Clark, & Porter, 1980). Here, the investigation of the instruction-assessment linkages views standardized achievement tests as one element in a large set of assessment methods whose impact on, and value for, students and teachers is evaluated in terms of learning and instructional factors.

As instructional factors have come into the limelight, the testing dialogue has taken the form of a debate, with the bulk of the test literature being a series of position papers citing little empirical data. This debate is being carried on predominantly by people outside the schools; the locus of the debate implicitly highlights the need to hear from teachers and those involved in daily classroom activities.

These position papers, in the main, focus on problems connected with the use of formal mandated achievement tests, either norm- or criterion-referenced. The National Educational Association (NEA) has taken the strongest position in the debate by asking that group standardized intelligence, aptitude, and achievement tests not be used to assess student potential or achievement until completion of a critical appraisal, review, and revision of current testing programs (McKenna, 1973). Others have criticized norm-referenced testing for its inefficiency, narrowness of foci, bias, invalidity, and unreliability (Broekhoff, 1978; Howe, 1978; Kahn, 1978; Klein, 1970; Perrone, 1978). The debate over appropriate measures continues with criterion-referenced testing being offered as an alternative because of its diagnostic, placement, and remediation information as well as its ease of interpretability in terms of specified performance standards (Broekhoff,

1978; Howe, 1978; Kahn, 1978, Klein, 1979; Nitko, 1971, Popham, 1978). This debate has helped to raise some important issues of test use but has not aided in their resolution. There is little empirical research available that can answer the questions that have arisen.

The focus of the research that does exist is predominantly on the use of standardized achievement tests. Goslin's (1967) study of testing at the elementary and secondary level represents the more recent comprehensive work and reports that in elementary schools, teachers use test results primarily to diagnose individual difficulties and to provide information to the student. However, he also reports that the teachers did not rely heavily on this source of information. Less than 20 percent had altered a course, and less than one third reported changing their methods as a result of tests. Stetz and Beck (1979) also conducted a nation-wide study of teachers' opinions of the use and usefulness of standardized tests; their study was conducted in conjunction with the standardization of the Metropolitan Achievement Tests. Teachers most frequently responded that they used test results for diagnosing strengths and weaknesses, measuring student growth, and evaluating individual students. The finding that 80 percent of the teachers reported making only some or little use of the data from standardized tests is similar to Goslin's conclusions.

The Royal Oak Study (Boyd et al., 1975) also supports the notion that teachers do not rely upon results of required or published tests for decision making. Although teachers reported variable use of results from the district-mandated testing program, there was little evidence that the testing program influenced school curriculum or classroom instruction. For the most

part, teachers felt that normed, standardized achievement tests were selected by administrators and imposed upon teachers, and did not furnish them with any new information. They felt that test results supplied information about students' skills that were already known by teachers and parents. Although a small number of teachers thought test results were useful, especially those from criterion-referenced tests, most felt that the tests given were not useful for planning instruction.

Literature on standardized testing has dominated the test use field of inquiry and there is evidence suggesting that such testing is on the increase. Kirkland (1971) reported that it was estimated in 1954 that more than 75 million standardized tests were taken by 25 million persons in educational institutions. In his study on teachers and testing, Goslin (1967) reported that in 1961 over 100 million ability tests had been given. Although the exact amount of standardized testing is not presently known, the National School Board Association (1977) polled more than 1,000 school board members and discovered that 75 percent reported the existence of district-wide testing programs which test every student or a large sample of students periodically, usually annually. From their survey, the NSBA concluded that district-wide standardized achievement testing affects over 80 percent of American school children. While these reports indicate there is some information about standardized achievement tests, little is known about the amount of other testing that takes place.

In addition to the standardized testing taking place in the majority of the districts in the United States, the minimum competency testing movement has swept the nation in a three year period. Gorth (1979) reported that 37

states had taken some kind of action to mandate the setting of minimum competency standards for elementary and secondary students. It is now reported that over 40 states have initiated minimum competency standards at various grade levels (Pipho, 1980).

Minimum competency testing has aldo received a good deal of attention in the testing debate. Proponents of competency-based testing such as Forbes (1978), Taylor (1978), and Reilly (1977), see it as a way of facilitating the reorganization of objectives so that they are sequential and appropriate. They also claim that minimum competency testing would aid in setting common public standards and that it would serve as a guide to identifying learning needs. The opponents of such testing are concerned that the minimum requirements are set too low and argue that minimum competency testing assumes that the minimums are known, credible, and measurable (Pipho, 1978; Reilly, 1977; Tierney, 1978; Wide, 1979).

Although much has been written about minimum competency issues, there has yet to be any report of the actual uses or extent of the use of competency-based tests. The National Evaluations Systems (1978) project has been given a grant to examine the minimum competency testing movement, but the project is only now underway. This project intends to survey competency test use and determine its impact on the schools.

In the literature, teacher-made assessments, curriculum-embedded tests, and district-constructed tests have been inspected much less closely than the formal measures discussed above. While the aforementioned figures suggest the magnitude of the more formal testing that is being done, virtually nothing is known about the amount of testing taking place using other types of assessments.

There is some evidence that curriculum-embedded tests and teacher-made tests matter in the course of instructional decision-making. In a CSE study of a small sample of California schools, Yeh (1978) reports that 55 percent of the teachers surveyed indicated that they regularly construct their own tests, which is an indication that some tests, if not published ones, are viewed by teachers as useful for some purposes. She also found that for assessing student progress, teachers tend to rely on more informal mechanisms such as observation and interactions with students than on the results of any tests. Others have only suggested teacher-made assessments as alternatives to standardized testing and offer techniques for teachers to help them make and use tests appropriately (Baron, 1958; Bauerfeind, 1963; Gorow, 1966; Traxler, 1963; Quinto, 1977).

The literature on curriculum-embedded tests is equally scant. David (1979) made some of the only references to curriculum-embedded testing in a study of the uses of federally mandated Title I evaluations. She interviewed school personnel, students, and parents and reported, among other findings, that the respondents felt that results of curriculum-embedded and other skill tests were more important than standardized tests.

Teacher observation and the subsequent judgments they lead to also clearly matter when a teacher is making instructional decisions. Leiter (1974), Mehan and Shumsky, (1973, reported in Mehan, 1974) and Bremme, Facchina, Kronish, and Wenger (1974) all point to the fact that teachers' personal assessments of students strongly influence their placement decisions.

Thus, evidence indicates that "other" assessments besides standardized tests deserve attention. The current information focuses on norm- and

criterion-referenced tests with some emphasis on minimum competency testing. Since literature on the other evaluative processis is lacking, there is a great need to look at various types of assessments to determine the purposes they serve school personnel. Tests are apparently used for diagnostic, placement, grouping, and evaluation purposes (Angel, 1968; Broekhoff, 1978; Carducci-Bolchazy, 1978; Nitko, 1971; Stetz, 1978; Wolek, 1972) but the specific tests used for these purposes are not known.

Factors Related to Use of Tests and Test Results

The kinds of contextual factors which influence testing and the use of test results are just beginning to be appreciated. The questions of how tests can contribute to teacher understanding of the individual pupil, and how to plan a testing program, select tests, and analyze and interpret test results for instructional improvement have been apparent for a number of years (Traxler, 1953). Teacher training, experience, and attitudes toward tests and testing seem to be key factors related to test use. Other factors that appear to matter are test-taker characteristics and the instructional options that are available to the teacher.

Teacher training and experience. Concern exists about the level of teacher training in testing. Most authors on the subject have recognized the need to improve teacher knowledge of tests and testing, to increase their involvement in the testing endeavor, and to facilitate their use of test results. Hastings, Runkel, and Danrin (1971) described a study in which the attitudes and perceptions of test users can be improved as a result of training. Ebel (1967) has also called for inservice training workshops to improve teacher competence in tests and testing. The literature does not appear to reflect any great follow-up to such suggestions.

Issues related to teacher knowledge and the use of tests abound and have been discussed by Boyd, McKenna, Stake and Yachinsky in their (1975) analysis of the Royal Oaks (Michigan) school testing practices. Among the problems impinging on test use were: minimal teacher involvement in the testing program; the purposes of the testing not being sufficiently conveyed to teachers; the content of the tests not always reflecting the goals of instruction and the background of students; training in administration of the tests and student preparation fortaking them being inadequate; too much staff time used in filing and storing test results; test score interpretation not being provided for; and minimal use of test reuslts because of teacher feelings of test irrelevancy or inadequacy.

Ebel (1967) raises somewhat similar concerns in his identification of errors commonly found in teacher constructed tests. Among the problems he discusses in teacher constructed tests are the following: too heavy reliance on subjective evaluation; leaving testing too late to be of instructional use; and developing tests that do not sufficiently sample student knowledge and ability in a given curricular area. More specifically, he cites development of trivial and ambiguous items; lack of teacher knowledge of the measurement errors to which tests are subject, and failure to test the effectiveness of their tests by statistical analysis of results.

Similarly, Leiter (1976) agrees about the quality of teacher constructed tests, and suggests that the background most teachers have in tests and testing will lead to the development of unreliable tests. Hastings et al. (1961) also agree that test use depends on knowledge of tests and their interpretation. This belief is implicitly seen in a number of works

(e.g., Gorow, 1966) which attempt to provide teachers with information on how to design teacher-made test items, and how to improve tests through analysis of test results.

Evidence confirms that teachers seem poorly prepared in the area of testing. Goslin (1967) provides one of the first extensive treatments of contextual factors surrounding teacher test practices. He asked teachers to respond to a variety of questionnaire items concerning: (1) familiarity with, and experience in, administering and scoring standardized tests; (2) opinions about the accuracy, fairness, and usefulness of these tests; (3) actual uses of tests, including reporting of scores to pupils and parents; and (4) teacher practices with respect to preparing students for taking standardized tests specifically.

Among Goslin's findings were that: (1) less than 40 percent of all teachers have had minimal formal training (one course) in test and measurement techniques, yet large numbers of teachers, especially in the elementary grades, are responsible for administering standardized achievement tests; (2) teachers tend to view standardized tests as relatively accurate measures of student achievement and to see the abilities measured by these tests as important determinants of academic success, but (3) teachers indicated a rather low degree of test use in terms of grading and advising pupils and in providing them with feedback (the higher the degree of teacher training in testing, the higher the use of test scores); (4) "coaching" for tests has not yet become a major activity of teachers, although many teachers do try to prepare students for tests. Goslin concludes that teachers will continue to be heavily involved in standardized testing. Among his recommendations are further consideration of the issue of teacher training in

tests and testing, and clearer school policy concerning the role of teachers in standardized test acministration and use.

There is still other literature that cites the lack of teacher knowledge of tests and testing. For example, Yeh (1978) reported that only 50 percent of the teachers sampled were able to interpret correctly two standard scores commonly used in reporting standardized achievement results (percentile ranks and grade equivalents). She concluded from this that teachers need more knowledge about testing and states that:

> Given these findings about teachers' knowledge and the fact that teachers indicated they wanted more training on how to use and construct criterion-referenced tests, it may be that teachers need more training before any potential value of the test is realized (Yeh, 1978, p. 42).

Yeh also determined in her study (1978) that more experienced teachers were more likely to use, and were most positive about standardized tests, than less experienced teachers. She hypothesizes that this phenomenon is a result of the recentness of the criterion-referenced testing movement, i.e., less experienced teachers had been trained in criterion-referenced testing and the limits of norm-referenced testing, while such exposure was not a part of more experienced teachers' initial training. Cramer and Slakter (1968) also support the relationship between teacher familiarity with tests and use of results. Teachers who were well informed about district testing programs and those who served on test advisory committees had a more positive attitude toward testing than teachers who did not.

Rudman and his colleagues at Michigan State University's Institute for Research on Teaching present findings which, in large measure, reflect all of the previous issues raised (Rudman, Kelly, Wanous, Mehrens, Clark,

& Porter, 1980). In a review of the literature dealing with the integration of assessment and instruction, Rudman et al., cite many works demonstrating that:

- teachers are prepared neither to construct their own tests nor to to interpret the results of standardized tests
- teachers are not helped either by the simple score many tests provide, or by the ponderous information yielded by overly-detailed listings of behavioral objectives
- many teachers cannot interpret commonly used standard scores
- there is an apparent need to help teachers in test construction, test administration, and test interpretation.

The literature on teacher preservice training and certification and teacher inservice training explains much about the lack of teacher knowledge of and attitude toward tests and testing; indeed, review of this literature makes it difficult to fault teachers for lack of knoeledge or for misconceptions about testing. For example, Woellner (1979) provides a national picture of teacher preservice training and certification. On a state-by-state basis, the certification requirements demand virtually nothing of teachers in terms of formal course work in testing (see Goslin, 1967, p. 127). For some school-based services staff, such as school psychologist or psychometrist, there are minimal requirements for course work in individual and group assessment. For some administrative/specialist credentials, an additional year may be required, but there does not appear to be much in the way of formal testing course work even for these positions.

In the vast majority of states, there are teacher requirements for cognitive course work in such possible test-related topics as the social or behavioral sciences or general psychology, but the extent to which these general courses may provide concrete information on tests and testing is probably quite small. Even in the descriptions of the professional course work for teachers, virtually no formal requirements for courses in testing appear; references to such concepts as testing for diagnostic and prescriptive purposes are non-existent. A few states (e.g., Indiana, Oklahoma, Tennessee) mentions, as part of professional course work, training in "evaluation of learning" or "measurement and evaluation." In each of these cases, the total number of semester hours required for all professional training is quite small so that the amount of time spent specifically on tests and testing is likely to be very limited.

In terms of inservice training, recent work again appears to make no direct statement of need to provide teachers with training in testing. Such work (Adam, 1975; Johnston, 1971; and Harris, 1980) generally considers the design, conduct, and assessment of successful inservice operations; it does not prescribe specific components in tests and testing as part of these programs.

It appears that such training is crucial to teacher knowledge of and use of tests. Whether one is discussing standardized test selection, teacher development of other forms of achievement tests, or alternative approaches to student assessment, teachers' knowledge and understanding of the particular bind of measure to be administered, their commitment to use of that measure, and the availability of linkages between the measure and instructional imporvement will be critical (Goslin, 1967).

Teacher attitudes. Teacher attitudes also have been found to play an important role in use of tests. Stetz and Beck (1978) employed survey questionnaires to assess teachers' opinions of achievement tests. The purpose of the study was not to determine teacher uses of test data, but to acquire empirical data about the widely held notion that standardized achievement tests are unpopular and disparagingly regarded by teachers. This assumption has frequently been presented as a reason for teachers' non-use of test data. Some 3300 teachers responded to the questionnaire which was comprised of eleven items, four of which can be construed as indicative of teachers' willingness to use test data.

The results showed that 55 percent of the total sample were neutral and 37 percent interested in using test data. It might be expected that teachers' reported lack of interest would be partly because of their feelings about test quality. Stetz and Beck's findings concur with this as they determined that 69 percent of the respondents felt neutral about the helpfulness of test data compared to 24 percent reporting that standardized tests were helpful. Other results showed that 64 percent fell within the neutral range on usefulness and 26 percent useful; 72 percent were neutral on test validity; only 16 percent considered tests valid.

Further empirical support of non use and low impact of test data and testing upon instruction comes from the "Irish study," which investigated the effects of standardized achievement tests on teaching practices and expectancies (Airasian, Kellaghan, Madaus, & Pedulla, 1977; Airasian, 1979). This study provided two sets of data. In the first reported results,

researchers compared three groups of teachers: those who received feedback on students' test performance on the standardized tests of reading and math achievement; teachers whose students were tested but who did not receive test data; and teachers whose students were not exposed to the testing. In the 1977 report, researchers collected teachers' ratings of their students' relative class standing in math and English before the November testing and in May (after the test results had been fed back to the first group of teachers). Correlations between initial ratings and test scores indicated that teachers' perceptions of their students were in accord with test outcomes.

The 1979 report of the study includes results from the same population on a questionnaire about perceptions, beliefs, and practices related to standardized intelligence and achievement tests, public exams, and classroom tests. Results were obtained for effects of exposure over time on perceptions of test accuracy, relative weight of data sources for instructional and selection decisions, including specific reports of weights for grouping decisions, and attitudinal factors relevant to use.

After four years, teachers who had been exposed without feedback to standardized achievement tests and teachers who had been receiving feedback of test scores showed a significantly greater rating in perceived accuracy of those tests. No such change in ratings was found for public exams and classroom tests. For teachers receiving testing information, further, accuracy ratings for standardized tests approached those given to classroom tests.

Despite the differences in perceived accuracy of tests among teachers in the three groups, there were no significant differences in the weights these teachers accorded to standardized test information for making instructional or selection decisions. Teachers' recommendations, however, were initially more highly rated across all teacher groups and did not differ significantly among the groups after the four years of differential exposure to standardized tests.

The impact of various data sources, including standardized achievement tests, was examined with regard to teachers' frequency of grouping students and the criteria used for grouping. From a list of ten possible sources, only the teacher recommendation criterion obtained significance for teachers in the no exposure/no feedback group; these teachers reported an increase in use of teacher recommendations as a grouping criterion. Availability of test data, then, did not alter grouping practices or criteria. Frequency of grouping (proportion of teachers in a treatment group reporting instructional grouping) did not relate to the availability of test data; no difference was found among the teacher groups.

Analysis of teacher attitudes toward standardized achievement tests revealed two dimensions accounting for response patterns: relevance to classroom instruction and decision-making; and expectancy effects from test data upon instruction. No difference was found in perceived instructional relevance of tests (slightly positive) among the teacher groups. However, effects of test data upon teachers' expectancies did differ significantly for teachers who received test data feedback. These teachers voiced greater disagreement with statements of negative impact of test data on

expectancies. Thus, availability of test scores did not affect the slightly positive perceptions of teachers about theclassroom relevance of tests. However, having been exposed to testing with feedback of results, teachers were less concerned than their counterparts in the other two groups about the impact of such data upon perceptions of pupils' abilities.

Another study concerned with teachers' attitudes and perceptions with regard to testing and test use was done by Salmon-Cox (1980). In the study, she employed open-ended interviews and extended classroom and non-classroom observations to explore the use of tests by thirty-five teachers in three suburban elementary schools. Despite differences among schools in tracking policy, socioeconomic status, and team-teaching characteristics, similarities were found in the teachers' general operating context and orientation with regard to testing and test use:

(1) with or without a lot of formally communicated information, many teachers reach judgments about their students early in the school year;

(2) greater emphasis is placed on social goals than on cognitive, and a "whole child" perspective dominates;

(3) home background characteristics are more salient in the classroom than are ability differences; and

(4) observation of students is the most frequently employed and highly trusted method for monitoring student progress (p. 13).

Teachers in all three schools reported that their principal use of test data is to confirm or verify their own judgments or to supplement their own information, viewing the test data as diagnostic or as a small part of

the larger general assessment of students. As in the Airasian study (1977), where discrepancies arise between test data and teacher ratings/expectancies, teachers tended to give the child the benefit of the doubt. Teachers questioned the accuracy of test scores when the test scores were lower than personal judgments of student ability, and they questioned their own judgments when these judgments were lower than the students' test results. In her summary of findings, Salmon-Cox emphasizes that low level use of test data is matched by low levels of concern for or negativity toward standardized and mandated testing.

Thus, teacher attitudes toward testing do seem to matter in the consideration of factors that influence test results. All of the studies mentioned included information about standardized achievement testing. As of yet, there is no evidence about how teacher attitudes toward other types of tests affect the use of those assessments.

Organization of instruction and related variable. The effect of the actual testing environment on test use is only beginning to emerge. Evidence suggests that characteristics of the test-takers and the instructional environment need to be explored. For example, Yeh (1978) found that the socioeconomic status and the average third grade achievement percentile of a school affected how tests were used. She also found that teachers with paid aides and teachers in team-taught classrooms tended to give a larger portion of tests that were locally developed by someone other than themselves (i.e., district tests) than did teachers without aides and teachers in self-contained classrooms.

Test takers' characteristics that were likely to influence test use included, for example, ethnicity, special developmental or learning problems, linguistic differences, age, socioeconomic status and other background variables, and test wiseness. (See Goslin, 1967). Meanwhile, instructional environment factors include personnel resources (such as teacher aides or resource teachers), instructional alternatives (such as pull-out programs or team-teaching situations), teachers' information, instructional practices, among others.

These factors have been considered in research on teachers' instructional decision-making or in studies of the social or organizational qualities of the classroom (see Lortie, 1965). The investigation of these variables as factors affecting teachers' use of tests and test data is minimal.

Other characteristics which conceivable influence teachers in their use of tests and which need to be considered in investigation of test use, are variables reflecting community and district issues. District issues include size, fiscal resources, linkages at the district level between curriculum and assessment, minimum competency testing, amount of centralization or decentralization in the district, and timeliness of test data feedback. In the community, parent involvement, accountability pressures, and news media coverage of test scores are possible influences on the nature and amount of testing, but they have yet to be researched.

Costs of Testing

We know very little about the costs of testing. A recent estimate (EDC News, 1977) was that in 1976, over 40 million elementary school children

standardized testing at a cost of well over a quarter of a billion dollars. This figure obviously does not include all the other types of tests that are typically administered to students.

Lyon (1978) reported that evaluation unit budgets at district offices range from $2,000 to $4,000,000, representing yearly per pupil expenditures ranging from $.10 to $90. The profile of expenditures for the average evaluation office revealed that 41 percent was spent on evaluation and 27 percent on testing (exclusive of research, development, and evaluation). Since Lyon also found that most evaluation activities involved achievement testing, the amount spent of testing is probably closer to 60 to 70 percent of the total budget.

Anderson (1977) looked at the costs of implementing minimal competency requirements and classified potential costs into four categories: (1) legislation, (2) implementation, (3) excess burdens thrust upon schools in terms of new programs, and (4) elimination of desirable elements from the curriculum. Of these four categories, he is only able to give cost estimates of test development, test administration, bureaucracy, and compensatory programs. To make testing programs cost-effective, Anderson recommends the following: (1) the state rather than local agencies should administer the minimum competency programs, (2) testing should not focus on the individual student, (3) cooperatives should be established to share tests, (4) test results should be reported publicly, (5) school districts should be rewarded for high performance, (6) financial inequities among school districts should be reduced, and (7) federally run testing should be discouraged.

The figures quoted above are not the sum total of fiscal costs associated with testing, but they do give some indication of its magnitude. There are also other costs of testing that need to be considered; e.g., costs in school and learning time, costs in staff time and energy to administer and review the tests, and psychological costs to teachers and students. Little information is available about these types of costs, and the little information that is available concerns teacher and student attitudes.

Stetz and Beck (1979) asked teachers to respond on a questionnaire to semantic differential scales, e.g., helpful-harmful, unbiased-biased, calm-anzious, and supportive-antagonistic. In general, teachers responded in the neutral range on each of these scales, with less than 10 percent indicating the negative extreme. Student opinions were also queried. At the K - 4 levels, a majority of students felt at least somewhat positive toward the test, although 67 percent indicated that they were nervous about taking the test. At higher grade levles (5 - 12), only 26 percent of the students felt positive about the test, while 27 percent reported feeling negative about the test. In addition, 30 percent reported getting nervous before taking teacher-made tests. One might expect attitudes and responses to be even more negative in the context of minimum competency tests, where the consequences of the test are more serious.

A study by Sharp (1966) also sheds light on some of the psychological costs of testing to teachers and students. Sharp mailed an opinion survey about standardized testing to 300 elementary and secondary teachers in Florida and interviewed twenty-five others by telephone. Eighty-four percent

of the teachers reported that heavy emphasis on tests affects teachers' mental health, and there was an evenly mixed reaction to the question of whether emphasis on testing caused competetiveness in the classroom.

The question of whether test scores affect a student's self-concept has also been raised. Kirkland (1971) pointed out that the effect of receiving information about one's abilities will depend on a variety of factors, including the legitimacy of the source of the information, the perceived accuracy of the test, the degree to which the information confirms one's own estimate, and the extent to which the information is threatening or rewarding. Test scores have potentially great impact where an individual's self-concept is at considerable variance with the record of performance on the test, where rationalizations of poor performance are unavailable, or where the test score is substantially high than one's own estimate. Under such conditions, one can expect a shift to affect the individual's aspiration level, motivation to achieve, and personal decisions about the future. However, data from a national sample (Kirkland, 1971) indicated that test scores are of relatively minor importance in shaping one's self-estimate of ability in comparison with school grades, comments made by peers and parents, and a student's relationship with his/her teachers. In addition, Kirkland reported that a majority of parents surveyed felt that their lives had been influenced by test results.

As indicated previously, information on any of the aforementioned issues is scant, and a need for understanding both the fiscal and psychological costs of testing is evident. The Test Use Project intends to explore these issues in depth in the second phase of the project when an

extensive field work methodology will better serve to answer the questions at hand.

Summary

In summary, the literature reviewed here provides some information pertinent to the project's conceptual framework. It was found, however, that the bulk of the information concerns norm-referenced, standardized tests, with much of this concentrating on the ongoing controversy concerning use of these tests. Other evidence suggests that tests of many types are being administered and the results are being utilized. To what extent this is occurring is not specifically known.

There are a number of areas concerning teachers and testing for which there is no information. One important area is the impact and use of test data from minimum competency and criterion-referenced tests. In particular, teachers have expressed concern that minimum competency tests will come to be used to evaluate the effectiveness of their instruction.* The potential effects of the minimum competency and criterion-referenced testing movements, as separate and distinct from the effects of standardized achievement tests, need to be considered.

The impact of other testing must also be considered. In-class assessments made by individual teachers have yet to be examined in depth. How these and other assessments are linked with teachers' instructional decision

---

*This concern was raised by teachers participating in the Test Use Project's exploratory fieldwork. This fieldwork was another planning activity influencing the design of the project's phase I national survey, and will be described in a future report.

making processes, and how they affect classroom organization and time allocations to other activities, are areas which should be explored. Teachers place greater reliance on, and have more confidence in, the results of their own judgments of students' performance, but little is known about the kinds of activities that give voice to this information about the extent to which they are used.

The settings and factors which affect the use of tests and their results is yet another uniformed area. The literature emphasizes training, experience, and attitudes as having a major role in teacher decision-making. But what of the instructional options, the practical needs, and the range of social behaviors of students that a teacher contends with daily? All of this needs to be taken into consideration when determining if, when, why, and how teachers are utilizing tests.

The literature suggests a great need for:

- broad-based information on the use and nature of the use of tests, especially those other than standardized tests;
- an examination of the following factors and how they affect test use
  - teacher training
  - teacher experience
  - teacher attitudes
  - instructional options
  - environmental factors
- an investigation of the cost-benefits of testing and test results

References

Adam, E. (Ed.). In-service education and teachers' centres. New York: Pergamon Press, 1975.

Airasian, P.W. The effects of standardized testing and test information on teachers' perceptions and practices. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, California, 1979.

Airasian, P.W., Kellaghan, T., Madaus, G.F. & Pedulla, J.J. Proportion and direction of teacher rating changes of pupils' attributable to standardized test information. Journal of Educational Psychology, 1977, 69(6), 702-709.

Anderson, S.B. Issues related to test use. Paper presented at the National Council on Measurement in Education, New York: 1977.

Angel, J.L. National, state, and other external testing programs. Review of Educational Research, 1968, 38(1), 85-91.

Baron, D. Evaluation techniques for classroom teachers. New York: McGraw-Hill Book Co. Inc., 1958.

Bauerfeind, R.H. Building a school testing program. Boston: Boughton Mifflin, 1963.

Borko, H. An examination of some factors contributing to teachers' preinstructional classroom organization and management decisions. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Boyd, J., McKenna, B.H., Stake, R.E., & Yachinsky, J. A study of testing practices in the Royal Oak (Michigan) public schools. Royal Oak, Michigan: Royal Oak City School District, 1975. (ERIC Document Reproduction Service No. ED 117 161)

Bremme, D.W., Facchina, M., Kronish, R., & Wenger, M. The culture and identity of teacher-student teams in a junior high school. Unpublished manuscript, Harvard University, 1974.

Broekhoff, M. Criterion-referenced evaluation as an alternative. English Journal, 1978, 67(7), 32-37.

Carducci-Bolchazy, M. A survey of the use of reading readiness tests. Reading Horizons, 1978, 18(3), 209-212.

Cramer, S., & Slakter, M. A scale to assess attitudes toward aptitude testing. Measurement and Evaluation in Guidance, 1968, 1(2), 96-102.

David, J.L. Local uses of Title I evaluations. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, April 1979.

Ebel, R.L. Improving the competence of teachers in educational measurement. In J. Flynn and H. Garber (Eds.), Assessing behavior: Readings in educational and psychological measurement. Reading, Mass.: Addison-Wesley, 1967, 171-182.

EDC News. Standardized Testing in America, Spring 1977.

Forbes, J.E. Some thought on "minimal competencies." The Mathematics Teacher, 1978, 71(2), 94-101.

Gorow, F.F. Better classroom testing. San Francisco: Chandler Pub. Co., 1966.

Gorth, W.P., & Perkins, M.R. A study of minimum competency testing programs: A final report. Washington, D.C.: Office of Testing, Assessment and Evaluation, National Institute of Education, 1979.

Goslin, D.A. Teachers and testing. New York: The Russel Sage Foundation, 1967.

Harris, B.M. Improving staff performance through in-service education. Boston: Allyn and Bacon, 1980.

Hasting, J.T., Runkel, P.J., & Damrin, E.E. Effects on use of tests by teachers trained in a summer institute. Urbana: University of Illinois, Bureau of Educational Research, 1961.

Howe, H., II. Tests and schooling. Two papers presented at The National Conference on Achievement Testing, Washington, D.C., 1978.

Johnston, D.J. Teacher's in-service education. New York: Pergamon Press, 1971.

Kahn, A. Remarks to the National Conference on Achievement Testing and Basic Skills, Washington, D.C., March 1978.

Kirkland, M. The effects of tests on students and schools. Review of Educational Research, 1971, 41(4), 303-350.

Klein, S.P. Evaluating tests in terms of the information they provide. Evaluation Comment, 1970, 2(2), 1-6.

Leiter, K.C.W. Ad hocing in the schools: A study of placement practices in two kindergartens. In A.V. Cicourel (Ed.), language use and school and school performance. New York: Academic Press, 1974.

Lortie, D.C. School-teacher: A sociological study. Chicago: The University of Chicago Press, 1975.

Lyon, C. Evaluation and school districts. Preliminary results reported to the National Institute of Education. Los Angeles, California: Center for the Study of Evaluation, 1978.

McKenna, B. Task force and other reports. Interim report on the task force on testing. Washington, D.C.: National Education Association, 1973. (ERIC Document Reproduction Service No. ED 151 421)

Mehan, H. Ethnomethodology and Education. Proceedings of the Second Annual Conference of the Sociology of Education Association, 1974.

National Evaluation Systems. Proposal submitted to the National Institute of Education for the review and evaluation of minimum competency testing programs, 1978.

National School Boards Association. Standardized achievement testing (Research report 1977-1). Washington, D.C.: National School Boards Association, 1977. (ERIC Document Reproduction Service No. Ed 152 797)

Nitko, A.J. A model for criterion-referenced tests based on use. Paper presented at the Annual Meeting of the American Educational Research Association, New York: April 1971. (ERIC Document Reproduction Service No. Ed 152 797)

Perrone, V. Remarks to the National Conference on Achievement Testing and Basic Skills, Washington, D.C., March 1978.

Pipho, C. Minimum competency testing in 1978: A look at state standards. Phi Delta Kappan, 1978, 59(9), 585-588.

Pipho, C. News from the states. NCME Measurement News, 1980, 18(5), 19-23.

Popham, W.J. Practical criterion-referenced measures for intra-state evaluation. Educational Technology, 1978, 18(5), 19-23.

Quinto, F. Teacher-made tests--an alternative to standardized tests. Today's Education, 1977, 66(2), 52-53.

Reilly, W. Competency based testing: More than a bandwagon? The Bangor Daily News, August 29-September 3, 1977. As reprinted in The Ford Fellows in Education Journalism Report, 1977.

Resnick, L., & Resnick, D. The social functions of educational testing. A proposal submitted to the Carnegie Corporatiok of New York, 1978.

Rudman, H.C., Kelly, J.L., Wanous, D.S., Mehrens, W.A., Clark, C.M., & Porter, A.C. Integrating assessment with instruction. A review 1922-1980. East Lansing, Michigan: Institute for Research on Teaching, 1980.

Salmon-Cox, L. Teachers and tests: What's really happening. Paper presented at the Annual M-eting of the American Educational Research Association, Boxton, Massachusetts, April 1980.

Sharp, E. How teachers see testing. Educational Leadership, 1966, 24(2), 141-143.

Stetz, F.P. Providing relevant test data for decision-making purposes. Elementary School Journal, 1978, 78(3), 220-225.

Stetz, F., & Beck, M. Comments from the classroom: Teachers' and students' opinions of achievement tests. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, April 1979.

Taylor, R. The question of minimum competency as viewed from the schools. Mathematics Teacher, 1978, 71(2), 88-93.

Tierney, R.J. Con: Minimal competency testing won't do it. Journal of Reading, 1978, 22(1), 10-12.

Traxler, A.E. Introduction to testing and the use of test results in public schools. Westport, Conn.: Greenwood Press, 1953.

Wise, A. Why minimum competency testing will not improve education. Educational Leadership, 1979, 36(8), 546-549.

Woelner, E.H. Requirements for certification for elementary schools, secondary schools and junior colleges. (44th ed.). Chicago: The University of Chicago Press, 1979.

Wolok, R.S. Let's use tests for teaching - standardized test results can provide the basis for a program of instruction. Teacher, 1972, 90(2), 62-64, 179-181.

Yeh, J.P. Test use in schools. Los Angeles, California: Center for the Study of Evaluation, 1978.